Pergamon

0042-6989(95)00101-8

Vision Res. Vol. 35, No. 20, pp. 2927–2941, 1995
Copyright © 1995 Elsevier Science Ltd
Printed in Great Britain. All rights reserved
0042-6989/95 $9.50 + 0.00

Two-dimensional Constraints on Three-dimensional Structure from Motion Tasks*

RICHARD A. EAGLE,† ANDREW BLAKE‡

Received 6 October 1994; in revised form 15 March 1995

Can humans recover metric structure from motion sequences or, as has been claimed by Todd and Bressan [(1990) *Perception & Psychophysics*, *48*, 419–430], are they limited to recovering only relief structure? Two experiments were carried out to investigate this question. In a metric-structure task, the angular thresholds for discriminating two rotating bi-planar structures were ≈91 deg. By contrast, in a relief-structure task, the angular thresholds for discriminating a planar from a non-planar structure, both undergoing simple rotational motion, were only ≈11 deg. A computational model is proposed to examine the image motion sensitivity required to perform discriminations of both three-dimensional metric and relief structure from motion. When the experimental data were re-plotted in terms of this two-dimensional sensitivity, the thresholds were found to be the same for both tasks. This finding is related to the model's revelation that recovering metric structure from motion is inherently more noise-sensitive than is recovering relief structure from motion. The conclusion is that the differences in angular thresholds reflect the differing nature of the two tasks. There is no evidence that the visual processes themselves are preferentially sensitive to non-metric over metric structure from motion.

Structure from motion Two-dimensional thresholds Three-dimensional thresholds

INTRODUCTION

Many perceptual and psychophysical studies have shown that when the human visual system is presented with a display simulating general motion of a rigid body, a powerful impression of a three-dimensional (3-D) structure is evoked (e.g. Wallach & O'Connell, 1953; Gibson & Gibson, 1957; Rogers & Graham, 1979). Perhaps surprisingly, it is only recently that theorists have begun to analyse the associations between processing requirements and performance levels on structure from motion (SFM) tasks (e.g. Ullman, 1983; Grzywacz & Hildreth, 1987; Todd & Bressan, 1990; Koenderink & van Doorn, 1991). One reason for this is that geometrical formulations for computing SFM have only been developed in the last 15 yr or so (Ullman, 1979; Longuet-Higgins & Prazdny, 1980; Hoffman & Bennett, 1986; Koenderink & van Doorn, 1975, 1991).

Metric vs relief SFM

Ullman (1979) originally proved that, under orthographic projection, accurate knowledge of the image positions of four non-coplanar elements in three distinct frames was both necessary and sufficient to determine the 3-D structure of a rigid object up to an overall scaling and reflection about the line of sight. If only two frames are available to the viewer, the structure and motion of the elements are confounded. This fact is illustrated in Fig 1(a). Three elements (●), one of which is fixed at the origin, are shown in plan view rigidly rotating around a vertical axis. However, the figure makes clear that an alternative configuration of elements can give rise to the same image motion (○). In particular, for an instantaneous rotation, or when the image displacements are small enough to approximate a temporal derivative, an alternative structure whose depth is a factor of η greater than the actual structure but whose rotation rate is a factor of η smaller will give rise to the same image motion (e.g. Todd & Bressan, 1990; Ullman, 1983). In principle therefore, when the instantaneous rotation is undetermined so is the structure, up to a single, affine stretch along the line of sight. This infinite family of structures all share the same bas-relief (ratio of depths) and all yield the same pattern of image motion. Additional elements cannot reduce the ambiguity but the

*This work was presented in preliminary form at the ARVO 1994 conference held in Sarasota, Fla.

†Department of Experimental Psychology, University of Oxford, South Parks Road, Oxford OX1 3UD, England.

‡Department of Engineering Science, University of Oxford, Parks Road, Oxford OX1 3PJ, England.

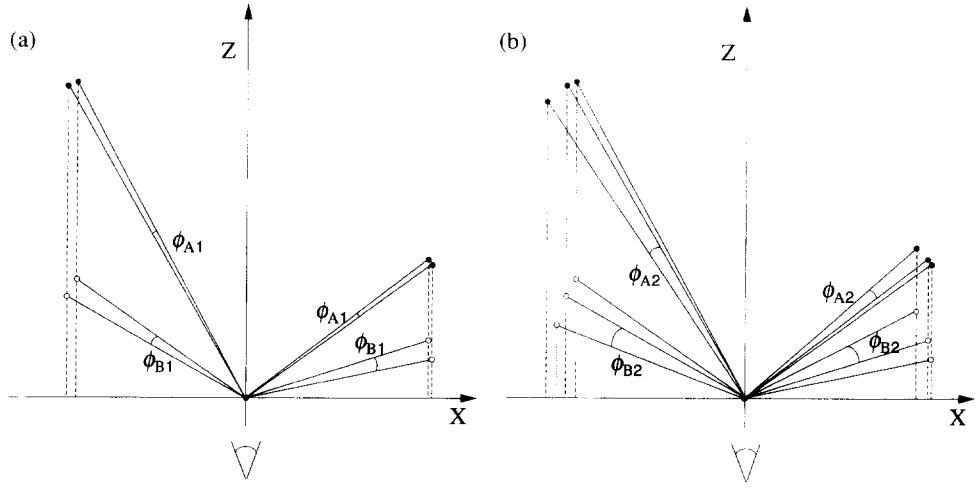

FIGURE 1. (a) Example of the structural ambiguity obtained with two-frame motion. Given the image motion produced by a three-element structure rotating around a vertical axis (shown here by the hashed projection lines) a family of metric structures are implicated. Two of these are shown here in this plan view, with structure A represented by ● and structure B by ○ (both structures share a stationary element fixed at the origin). Note that in order that the two structures elicit the same image motion, they are constrained to share approximately the same bas-relief (relative depth of elements) with the deeper structure having a proportionally smaller rotation angle ($\phi_{A1} < \phi_{B1}$). (b) Here, the two-frame case is augmented by a third view. For each member, N, of the family of possible structures considered above, the new image motion of, say, the right-hand element essentially fixes the rotation angle, ϕ_{N2} (not necessarily the same as ϕ_{N1}). Given this rotation angle, there is a predicted image motion for the left-hand element. However, for the two structures considered here, only A is compatible with this image motion. The predicted image motion for the left-hand element of B (shown by the dotted line) is smaller than the observed motion. In fact, out of the whole family of possible structures compatible with any two of the three frames, A is uniquely compatible with the image motion from all three.

addition of a third frame (second temporal derivative) leads to a unique solution of both the structure and motion, as illustrated in Fig. 1(b). A fourth element is not required in this case because the axis of rotation is fixed across the two displacements (Hoffman & Bennett, 1986).*

Adopting a different tack, Todd and Bressan (1990) and Koenderink and van Doorn (1991) have emphasized that even though metric properties cannot be determined from two-frame motion sequences, knowing the bas-relief structure of an object still allows many 3-D tasks to be performed. In fact, Todd and Bressan (1990) pointed out that in the past many of the SFM tasks that experimenters have employed (e.g. detecting non-rigidity or judging ordinal depth relations) required only the determination of relief structure. In an influential study of their own, they specifically designed experiments that required the recovery of Euclidean SFM and compared subjects' performance to that for tasks that required only relief or, in their term, affine SFM to be recovered.

In one Euclidean-structure task, observers had to judge whether the dihedral angle between two planes, defined by a set of lines rotating around a backward-slanting axis, was greater or less than 90 deg. They found that threshold performance was reached when the angle was ± 18 deg from 90 deg and increased only slightly as the number of distinct frames in the sequence was decreased from eight to two—an apparently surprising result, given Ullman's theorem. However, Todd and Bressan pointed out that if observers had assumed a rotation angle of between 0 and 90 deg for the structures (drawn on each trial from a rectangular distribution) then this threshold would be expected, even for the two-frame condition. In sum, Todd and Bressan found no evidence that subjects were able to derive either the metric structure or the rotation angle from the image motion, which led them to conclude that the human motion system is not capable of performing a full Euclidean analysis of the scene.

They also performed an experiment which required 3-D SFM to be computed but only up to an affine transformation. In this case, subjects were asked to judge which of two rotating four-line structures was planar. For the non-planar structure, the angle between the two planes formed by the four lines was systematically manipulated. Results showed that threshold performance was reached when the deviation from non-planarity was ± 2.6 deg, and again did not vary as the number of novel frames presented was decreased from eight to two. As a structure can be specified as planar or non-planar in a two-frame sequence, Todd and Bressan took their results to mean that humans can utilise first-order temporal relations for SFM recovery with high sensitivity. Their overall conclusion was that

*In the case of perspective projection, Longuet-Higgins and Prazdny (1980) have shown that Euclidean structure can be derived from just two frames of motion. However, the processing required to exploit the additional information available under perspective projection (in their case knowledge of the instantaneous velocity field and its first two spatial derivatives) is highly noise-sensitive: greatly different 3-D structures can give rise to very similar velocity fields (Ullman, 1983). Given this, Ullman (1983, 1984) has suggested that the human motion system is unlikely to rely on this information.

human observers process SFM only up to an affine transformation along the line of sight, but that metric structure cannot be recovered. Subsequent studies have built on this geometric framework to provide a more detailed picture of human performance on different types of SFM tasks and have produced results largely consistent with Todd and Bressan's (1990) conclusion (e.g. Braunstein, Hoffman & Pollick, 1990; Liter, Braunstein & Hoffman, 1994; Norman & Todd, 1993).

Required image motion sensitivity

An alternative explanation for Todd and Bressan's finding that recovery of metric SFM is poor is that the required sensitivity to image motion for the metric structures to be discriminated was too high. In order to investigate this possibility, it is necessary to express sensitivity to metric and relief SFM in terms of the underlying two-dimensional (2-D) motion thresholds. This paper describes a computational model of the sensitivity to image motion required for a metric and a relief SFM task along with two experiments designed both as a test of the model and as a gauge to human performance. Rather than use existing data, these new experiments were necessary for two reasons. First, it was important to discover whether performance for a metric-structure task *ever* rises above chance when unaided by confounding cues. Second, data from experiments on relief structure, such as Todd and Bressan's (1990), were inappropriate for analysis by the model proposed here since vital parameters affecting the model's output, such as the initial viewpoint of the structure, were randomised in the experiments.

To anticipate the empirical findings, it is shown that observers *can* compute metric SFM under such circumstances. Moreover, the combination of the empirical data along with results from the computational analyses suggest that the sensitivities to metric and relief SFM are equivalent when expressed in terms of the required thresholds for detecting changes in the retinal 2-D flow-field. The strong implication of these findings is that the human visual motion system is not specifically tuned to detecting relief SFM in preference to metric SFM. Rather, metric SFM tasks are inherently more noise-sensitive than are relief SFM tasks.

GENERAL METHODS

Apparatus

All stimuli were generated on a Silicon Graphics Crimson/VGX workstation. They were presented on a 19 in. Silicon Graphics monitor with a screen resolution of 1280 × 1024 pixels. Viewing distance was 86 cm in all experiments such that each pixel subtended 1 arc min at the eye. Subjects' responses were recorded via a mouse linked to the workstation.

Stimuli

All stimuli were composed of randomly-positioned dots, where each dot was specified by a 2 × 2 block of pixels. The stimuli were 5.33 deg high × 6.66 deg wide and were filled with a total of 200 dots plotted for each frame. Antialiasing, using 8-bit specification, was used to achieve sub-pixel plotting accuracy (e.g. Foley, van Dam, Feiner & Hughes, 1990). The simulated 3-D structures were comprised of two planes rigidly hinged together to form a constant dihedral angle coincident with the vertical rotation axis. A central vertical bright line was present at the location of the hinge. Figure 2 shows an example of a single stimulus frame. All projection was orthographic.

Procedure

For a particular trial, observers had to discriminate between two stimuli viewed in alternation. Of the two stimuli, one was a "standard" structure and the second was a "comparison" structure, which was selected at random from a set of five according to the method of constant stimuli (e.g. Engen, 1972). Observers were able to switch back and forth between stimuli at will until they felt able to make the required judgement, according to a two-alternative forced-choice procedure. Each comparison structure appeared in 10 trials so that a block contained 50 trials. Subjects completed four blocks of trials for each condition, and so each condition was tested 40 times.

Prior to the experiment proper, subjects performed two blocks of trials for each condition for which the resulting data were not analysed. This served both as practice and enabled an appropriate range of comparison structures to be prepared for the experiment proper. No feedback was given during any part of the practice sessions or experiment proper. Subjects viewed the stimuli in a dark room with an opaque patch over their weaker eye. One of the authors (RAE) and two naive subjects participated in the experiments.

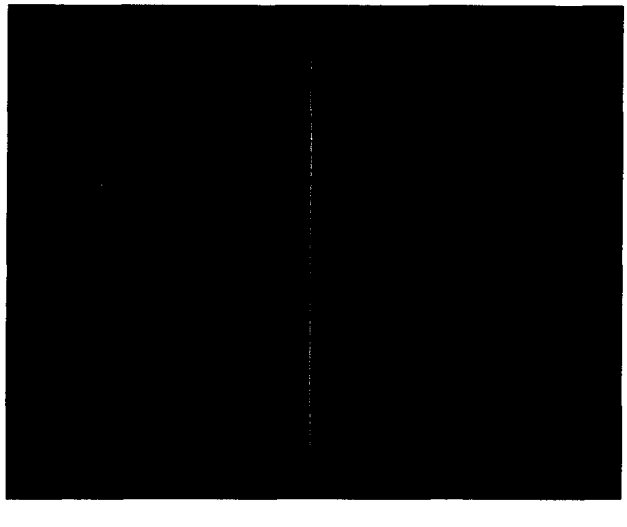

FIGURE 2. One frame from an example stimulus, used for the experiments. The vertical line specifies the location of the rotation axis as well as the hinge at which the two planar facets meet.

EXPERIMENT 1: SENSITIVITY TO METRIC SFM

The specification of an object's metric structure includes a description of the angles between its features in 3-D space. In order to investigate subjects' ability to recover such 3-D angles from motion sequences, a simple discrimination task was devised. The stimuli for this task, illustrated in Fig. 3, simulated rigid structures that appeared like "open-books" with a vertical hinge and that rotated around a vertical axis in three-frame motion (see also Hogervorst, Kappers & Koenderink (1993) who used similar stimuli for measuring structure from motion thresholds). Subjects saw two such structures on every trial and their task was to choose the structure with the smaller dihedral angle between its two planar facets (i.e. which book was more closed).

Liter *et al.* (1994) found that if two stimuli have similar simulated 3-D structures but different rotational magnitudes, subjects perceived the structure with greater rotation angle to have greater depth. They have argued that this was due to the fact that the structure with the faster rotation had a greater amount of relative image motion. In order to eliminate this potential biasing factor from the present experiment, the rotation angle of each structure was co-varied with its dihedral angle. The co-variation was arranged so that elements on different structures sharing the same image location in the first frame also shared the same image location in the final frame (see Fig. 3) i.e. so that the cumulative image displacements of such elements were identical across all structures. This meant that structures with greater depth rotated in smaller angular steps (see also Bradshaw, 1989, Chap. 5).

Methods

Three-frame condition

A crucial aspect of all stimuli, illustrated in Fig. 3, was that the structures were symmetrical around the line of sight in the second frame. The rotation angle was constant across displacements and was initially anti-clockwise from a plan view. The projected lengths of the left and right planar facets were equal in the second frame which meant that the total image motions of any two texture elements on the left and right facets equidistant in 3-D space from the rotation axis were equal in magnitude (but opposite in sign). The projected lengths of the left and right facets of all structures in frame one were 3.4 and 3.2 deg respectively and thus 3.2 and 3.4 deg in frame 3. (In practice, because the facets were only visible by a covering of random dots, these values represent upper bounds). This meant that while the 3-D lengths of the two planar facets were the same within a structure they differed markedly across structures.

The dihedral angle for the standard structure, ω_1, was fixed at 58 deg and the rotation angle per frame, ϕ_1, was 1 deg (i.e. the cumulative rotation angle for the structure was 2 deg). This meant that the angle of the right facet in frame one relative to the image plane, θ_1, was 60 deg, while that of the left facet was 62 deg. The initial angle of the right facet with the image plane for the five comparison structures, θ_2, was 8, 11, 14, 17 or 20 deg. The rotation angle of the comparison structure, ϕ_2, was set so that texture elements in the first frame that shared image location as elements on the standard structure also shared the same cumulative image displacement. The rotation angles used for the five comparison struc-

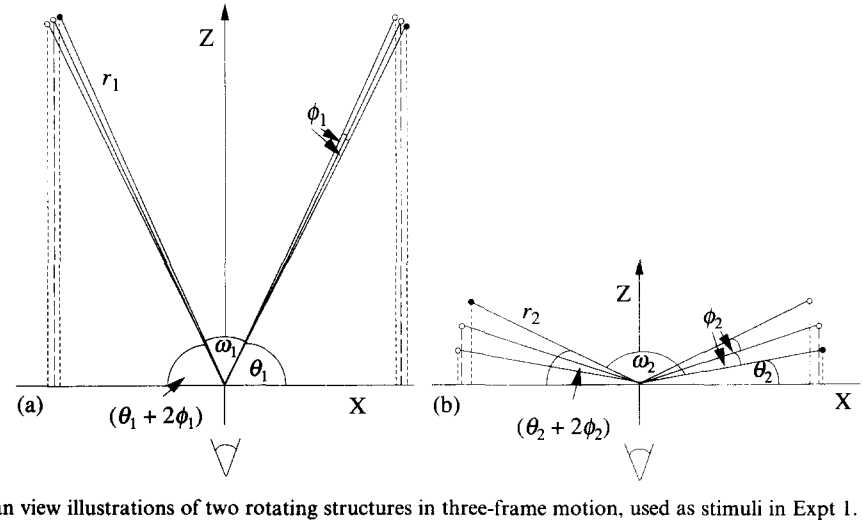

FIGURE 3. Plan view illustrations of two rotating structures in three-frame motion, used as stimuli in Expt 1. Each structure, shown separately in (a) and (b), is depicted by a series of three V-shapes which represent the two planar facets across three frames of a motion sequence. These planar facets were covered in random dots in the experiment. The ● represent the endpoints in the first frame, with the ○ showing both the second and third frames of an anti-clockwise sequence. Both structures are symmetrical about the line of sight in the second frame. The projected image motion is shown by the vertical hashed lines. The standard structure, shown in (a), has a smaller dihedral angle between its two facets ($\omega_1 = 58$ deg) than the comparison structure ($\omega_2 = 150.4$ deg), shown in (b). The 3-D length of the standards' facets is greater than that of the comparisons' but the rotation angle is proportionally smaller ($\phi_1 = 1$ deg, $\phi_2 = 6.8$ deg). The consequence of these settings is that the image positions of the endpoints in the first and last frames are identical across the two structures, as shown by the equal separation of the two outer projection lines in (a) and (b). These two structures can only be distinguished by their second-frame projections, illustrated here by the middle, long-dashed lines. These reveal that there is a larger difference in the image motion for the endpoints of the structure in (b) across the two displacements.

tures were 6.80, 5.92, 5.18, 4.56 and 4.04 deg. Finally, the dihedral angles of the five comparison structures, ω_2, can be calculated simply as $180 - 2(\theta_2 + \phi_2)$ and can be seen to be 150.4, 146.16, 141.64, 136.88 and 131.92 deg respectively. This range of rather large dihedral angles, compared with that of the standard, was chosen on the basis of observations from a pilot study.

Frame duration was 166.6 msec and there was no ISI.

Fifteen- and two-frame conditions

Two other conditions were investigated in addition to the three-frame case. Both employed the same structures as the three-frame condition but varied the temporal nature of the displays. In one, the stimuli were simply sampled at a higher temporal frequency, such that there were 14 displacements (15 frames) instead of just two, each displayed for 33.3 msec. Thus, the cumulative exposure duration for a single sweep, now from frame 1 to 15, remained 500 msec. As before, the structures were in continual oscillation.

While Ullman (1979) has shown that three frames of motion are necessary for metric structure to be computed, it is important to show empirically that subjects could not perform the task given only a two-frame sequence. In a third condition, this hypothesis was tested. Each frame was exposed for 166.6 msec, as in the three-frame condition, but now only the first two frames of this sequence were shown (again in continual oscillation). Because the motion of the left-hand facet in the clockwise swing was identical to the motion of the right-hand facet in the anti-clockwise swing, there was no need for an additional control test to be performed with the second two frames. The reason for not choosing the first and last frames as the stimulus for the two-frame condition was that these were identical for all structures and hence would have provided no information even for the two structures to be distinguished. Using only the first two frames has the effect of halving the total rotation angle, which might be thought, in and of itself, to make the task harder. In fact, because of the aforementioned symmetry between the first and second displacements, and that the first and third frames were identical across all stimuli, the extended rotation angle of the three-frame condition provided no benefit over the two-frame condition *per se*.

It is strictly invalid to refer to two-frame orthographically projected stimuli as simulating a unique metric structure, as discussed above. However, for the purposes of plotting the data, the two-frame stimuli are subsequently referred to as having a metric structure which is identical to that of the multi-frame stimuli.

Results and Discussion

The psychometric functions for the three conditions are shown for three subjects in Fig. 4 along with the mean data across all subjects.

The data from the three- and 15-frame conditions were fitted with Weibull functions (Weibull, 1951) of the form:

$$f(x) = 100 - 50\exp[-(x/A)^B], \quad (1)$$

where A is the angular difference at 81.6% correct performance and B is the slope. Thresholds for discriminating two structures were taken as the difference between the dihedral angles of the standard and comparison structures at which 75% of the judgements were correct. Averaged across subjects, these thresholds were 91.3 deg for the three-frame condition and 91.1 deg for the 15-frame condition ($\omega_2 = 149.3$ and 149.1 deg

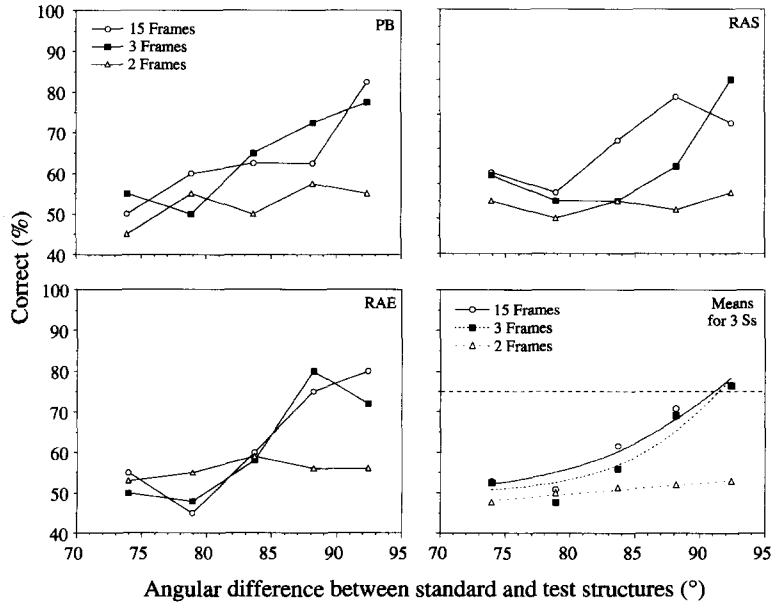

FIGURE 4. Psychometric functions for dihedral angle discrimination for three subjects performing three conditions. The abscissa plots the difference between the dihedral angles of the comparison and standard structure, the latter being fixed at 58 deg. See text for a description of the three conditions. For the mean data set, the three- and 15-frame condition slopes are fitted with a Weibull function [equation (1)] while the two-frame condition slope is fitted with a linear function.

respectively). Performance in the two-frame condition showed no improvement beyond 50% correct with an increasing angular difference between the structures. Taken together, these results provide strong evidence in favour of the visual system being able to integrate information across two displacements in order to determine metric structure. The fact that increasing frame numbers beyond three (higher temporal sampling) did not lead to decreased thresholds suggests that the higher-order information in these stimuli was not used by the visual system.

A threshold of 91 deg for discriminating two dihedral angles is perhaps too poor to be of use in many real-world tasks requiring an estimate of metric structure. There are, however, at least two ways in which sensitivity could be increased. First, it must be remembered that motion provided the only useful information about the simulated depth of structures in this study (all other depth cues present were consistent with a flat surface). One way of decreasing depth judgement thresholds might be to combine motion information with that from other cues, such as stereo (e.g. Johnston, Cumming & Landy, 1995). Second, even if no other depth information is available, it should be noted that the stimuli used in the present experiment did not contain large amounts of motion—the total rotation of the standard structure in Expt. 1 was only 2 deg, for instance. A larger rotation angle may well provide additional information to allow the visual system to perform finer discriminations.

EXPERIMENT 2: SENSITIVITY TO RELIEF SFM

This experiment was designed to investigate subjects' ability to discriminate structures on the basis of their 3-D relief. Subjects' task was to decide which of two structures was planar when both were viewed rotating about a vertical axis. The property of planarity is specified by a two-frame sequence, even for arbitrarily large discrete rotations.

A similar experiment was performed by Todd and Bressan (1990). They found that angular thresholds on their task were as low as 2–3 deg. Rather than use these data a new experiment was performed, with the stimulus parameters designed to match as many of those in Expt 1 as possible. The goal was to enable a direct comparison of sensitivity across relief and metric SFM tasks. As mentioned earlier, it is only strictly possible to measure 3-D angular thresholds when the stimuli are rendered in at least three frames of motion. With the two-frame sequences used in this experiment, the metric structure was undetermined and hence so was the absolute angular deviation from planarity. To overcome this, both a two-frame and a 10-frame condition were included where, as in Expt 1, the 10-frame condition was generated by higher temporal sampling of the same 3-D structure undergoing the same rotation. A three-frame condition was not included for this task as, unlike for the metric-structure task, a third frame would not have added any information significant for the task.

Methods

The spatial extent, dot density, and frame duration for the stimuli were identical to those used in the three-frame condition of Expt 1. Again, subjects were allowed to switch back and forth between two rotating structures until they felt able to make their response. The angle of the planar (standard) structure with respect to the image plane, α_1, was 60 deg and the rotation angle, ϕ_1, was 1 deg (see Fig. 5). The five non-planar comparison structures contained a rigid hinge at the vertical rotation axis, much like the stimuli in Expt 1, such that the

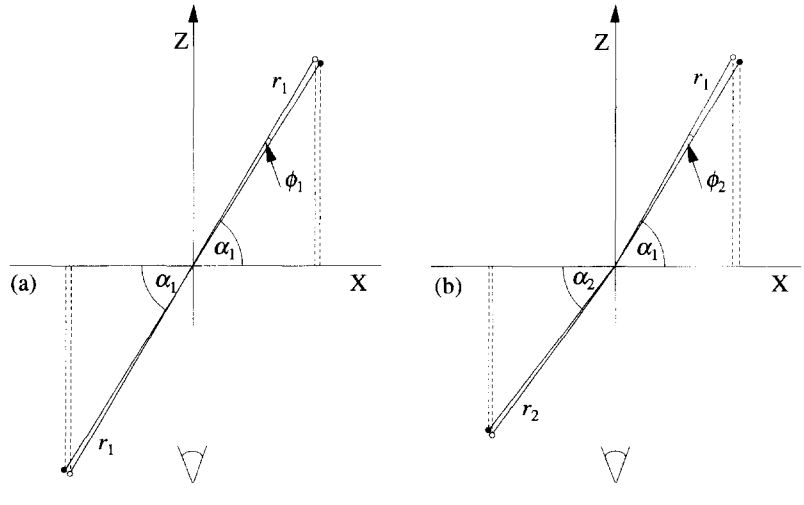

FIGURE 5. Plan view of two structures rotating in two-frame motion around a vertical axis. The ● depict the endpoints in frame 1 while the ○ represent frame 2. As in Fig. 3, the image motions of the endpoints are depicted by the dashed vertical lines. In (a) is shown a planar figure where the angles of both facets with the fronto-parallel, $\alpha_1 = 60$ deg. In (b) a non-planar figure is shown where $\alpha_1 = 60$ deg and $\alpha_2 = 52$ deg. The 3-D length of the non-planar structure's left-hand facet, r_2, has been chosen to ensure that the image position of the endpoint is identical to that belonging to the planar structure. ϕ_1 and ϕ_2 are the rotation angles for the planar and non-planar structures respectively. $\phi_1 = 1$ deg and $\phi_2 = 1.15$ deg—a value chosen to normalize the sum of the image displacements for the left- and right-hand endpoints across structures.

dihedral angle of the two planes was fixed over time. The angle of one of the planar facets with respect to fronto-parallel was 60 deg, while the angle of the other facet (randomly the left or right across trials) was variable. The exact slant of these facets, α_2, was chosen from within the range of 58–46 deg, depending on the subject's performance as gauged in a practice session. These can be re-stated in terms of a deviation from planarity as a range of 2–14 deg.

Using the geometry and notation shown in Fig. 5(a), the projected lengths of the two facets in frame 1 were held constant both within each structure and across structures by co-varying the 3-D length, r_2, of the facets with their initial slant. Analogously to Expt 1, an attempt was also made to normalize the total image displacements of elements symmetrical in the image about the rotation axis in frame one across all structures, although the significance of this control is not as great as in Expt 1. In order to do this, the rotation angle of the comparison structure was co-varied with the deviation from non-planarity, such that, e.g. when $\alpha_2 = 58$ deg, $\phi_2 = 1.04$ deg and when $\alpha_2 = 46$ deg, $\phi_2 = 1.22$ deg.

Results

The psychometric functions for the two-frame and 10-frame conditions are shown in Fig. 6. Thresholds were again taken as the deviation from planarity at which 75% of the responses were correct, following the data being fitted with a Weibull function. For the 10-frame condition, the mean threshold across subjects was 10.7 deg. Taking the metric structure of each two-frame stimulus to be the same as that of the corresponding ten frame stimulus, the mean threshold difference across subjects for this condition was 10.9 deg. This implies that the minimum number of frames necessary to perform the task (two) contained all of the information used by the visual system. A comparison of these data with those illustrated in Fig. 4 shows that these discrimination thresholds are a factor of about 8 smaller than those elicited by the same subjects when required to discriminate metric structures. This is consistent, at least qualitatively, with other empirical findings discussed in the Introduction (e.g. Todd & Bressan, 1990).

IMAGE-MOTION SENSITIVITY REQUIREMENTS FOR SFM RECOVERY

In this section an information analysis of the two SFM tasks used in Expts 1 and 2 is provided. It combines psychophysical findings on human low-level motion sensitivity with a geometric analysis of the consequences of this sensitivity for the computation of 3-D structure. The goal is to offer a unifying interpretation of the performance on metric- and relief-structure tasks.

Extracting information from the flow-field

For both tasks, it was found that higher temporal sampling did not lead to a reduction in thresholds. This indicates that the frames in between those used in the minimal-frame conditions (three-frame for metric SFM and two-frame for relief SFM) were not useful to the visual system. The implication is that SFM recovery is based on determining the image positions of spatial elements across discrete frames. An obvious limit to human performance, therefore, is set by the error on these positional measurements.

For both metric and relief SFM, two strategies are considered for processing this image information: (1) measurements of the image positions of elements across frames and (2) measurements of the image displacements

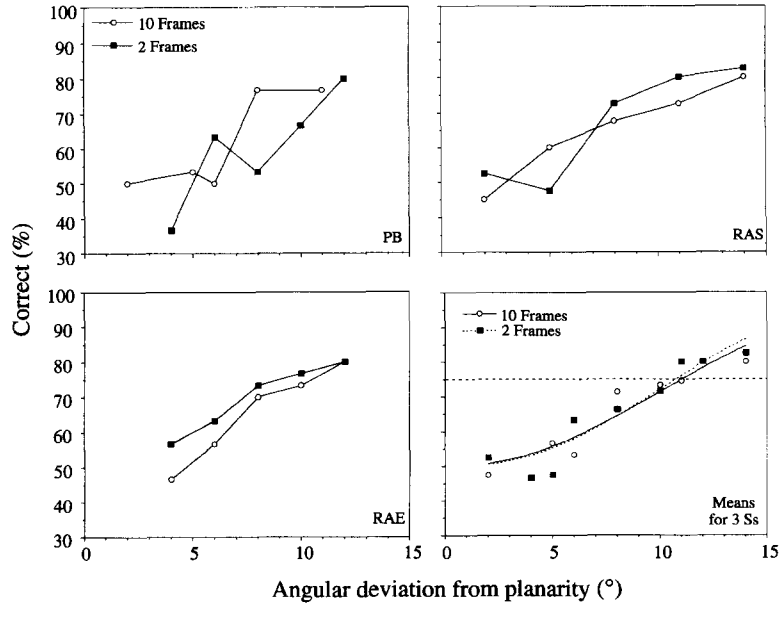

FIGURE 6. Psychometric functions for planarity discrimination for three subjects performing two conditions, along with the mean of their data. The abscissa plots the deviation from planarity of the non-planar structure, where the planar structure slants back from the fronto-parallel at 60 deg. The two-frame structures are assumed to have the same metric structure as the more densely-sampled 10-frame structures. For the mean data set, the slopes are fitted with a Weibull function [equation (1)].

of elements across frames. On the basis of known Weber's law thresholds for analogous 2-D tasks it is concluded that the latter technique more closely models human performance.

Metric SFM

One possible strategy that the visual system could use to determine the image positions of elements across frames would be to measure the locations of elements separately on every new frame, relative to the rotation axis. The thresholds observed in psychophysical experiments in which subjects were required to discriminate the lengths of two lines, or to judge the spatial separation of elements are directly relevant to the efficacy of this strategy. In such experiments, the data have been shown to follow Weber's law, with Weber fractions of around 0.03 under optimal conditions (Westheimer & McKee, 1979; Watt, 1987).* The data obtained in the present Expt 1 can be re-expressed in terms of such a Weber fraction to determine whether such a strategy would be capable of supporting the performance levels achieved.

For the two displacements, the changes in image positions of all elements on the right (left) facet of the standard structure were 0.0304 (0.0307) and 0.0307 (0.0304) of their projected distance to the rotation axis. Given the above thresholds for line length discrimination and separation judgements, this strategy might just have afforded discrimination of the three frames of a single stimulus, i.e. noting that there had been some change. However, for this strategy to be successful, a further requirement would be the discrimination of the image positions of elements on the standard and comparison structures. The difference in the positions of elements on the threshold comparison structure and standard structure that shared the same image positions in frames 1 and 3 was a factor of only 0.007 of the mean distance to the rotation axis. When this is compared to the known Weber fraction of 0.03 for such discriminations, it is clear that this strategy is incapable of supporting the 3-D thresholds obtained in Expt 1.

A second strategy that the visual system might employ would be to update positional information by measuring the image displacements of texture elements. Results of experiments by Snowden and Braddick (1991) and Werkhoven, Snippe and Toet (1992) are relevant in considering the psychological plausibility of this proposal. Snowden and Braddick (1991) showed subjects two random-dot patterns. The speed of one pattern was made to vary over time as a square-wave function, while the other was moved at a constant velocity, equal to the mean of the first pattern. The subjects' task was to decide which pattern had varying speed. The results show that at low temporal frequencies, discrimination thresholds can be expressed as a Weber fraction (difference in speed divided by mean speed) of around 0.3 at their lowest. Similar data have been recorded by Werkhoven *et al.* (1992) using isolated dot elements. These data can also be expressed in purely spatial terms. For this case, thresholds are a constant Weber fraction when the stimuli are characterised in terms of the change in image displacement over time as a proportion of the mean displacement.

The implications of these findings for the present three-frame metric SFM task are as follows. The image motion of all elements across the three frames can be expressed in terms of the change in displacement as a proportion of the mean displacement. For the standard structure, this formulation yields values of approximately zero (± 0.0097) for all elements. Therefore, the crucial 2-D processing required to support the 3-D discriminations would be to detect the displacement variation of elements on the comparison structure. For the threshold comparison structure, the changes in displacement of all elements divided by their mean displacement were ± 0.41. The fact that this value is greater than that the discrimination thresholds found by Snowden and Braddick (1991) and Werkhoven *et al.* (1992) shows that this second strategy, in which the image positions of elements are updated by measuring subsequent image displacements, could support the 3-D thresholds obtained in Expt 1. The subsequent analyses of the sensitivity to image motion required to compute metric SFM in this paper assume this strategy as their basis.

Geometric analysis of dihedral angle discrimination

Here, the consequences of errors on estimates of the image motion are considered in detail for the family of structures used in Expt 1 over a range of rotation angles. The change in image displacement as a proportion of the mean displacement is defined simply as

$$\Delta d/d_m = \frac{d_2 - d_1}{(d_1 + d_2)/2}, \tag{2}$$

where d_1 and d_2 are the two image displacements. Using the notation shown in Fig. 3(b), $\Delta d/d_m$ for a single element undergoing two equal rotational steps around a vertical axis is

$$\Delta d/d_m = \frac{[r_2 \cos(\theta_2 + \phi_2) - r_2 \cos(\theta_2 + 2\phi_2)] - [r_2 \cos\theta_2 - r_2 \cos(\theta_2 + \phi_2)]}{[r_2 \cos\theta_2 - r_2 \cos(\theta_2 + 2\phi_2)]/2}. \tag{3}$$

Using substitutions, factor formulae and double angle identities, (3) becomes

$$\Delta d/d_m = \frac{2 \tan(\phi_2/2)}{\tan(\theta_2 + \phi_2)}. \tag{4}$$

Thus, it can be seen that for small values of ϕ_2 (small rotation angles), $\Delta d/d_m$ approximates a cotangent

*All Weber fractions from the empirical literature are reported here in terms of the change in intensity divided by the *mean* intensity, although these have generally been converted from the original formulations in terms of the change in intensity divided by the *lower* intensity. The reason for this conversion is to make explicit the comparison with the present data which are most naturally expressed in the former terms. These changes are purely for ease of comparison and are not otherwise significant.

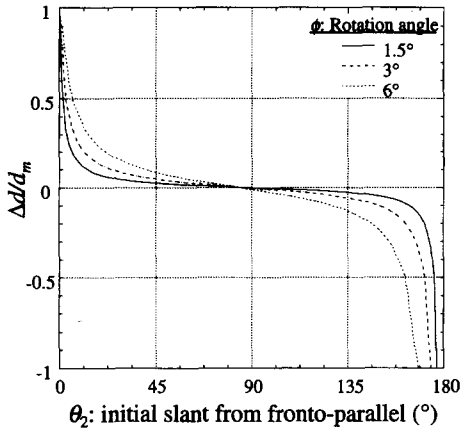

FIGURE 7. Values of $\Delta d/d_m$ for a single element rotating in two angular steps of ϕ_2, with initial slant θ_2. The curves are derived from equation (4) for three values of ϕ_2.

function, with the amplitude scaled by ϕ_2, as illustrated in Fig. 7. Note also that r_2 cancels out, showing that $\Delta d/d_m$ has a constant value for all elements lying on a plane that rotates about an axis formed by its intersection with the image plane. Figure 7 shows several members of the family defined by (4). While this graph specifies the initial slant and rotational angle for only a single element, it is important to note that the structures used in Expt 1 were symmetrical around the line of sight in the second frame. This means that each point on the curve in Fig. 7 describes the magnitude $\Delta d/d_m$ for *all* elements on *both* planar facets of a particular structure, although the sign will be reserved for elements lying on the left-hand facet.

These values of $\Delta d/d_m$ for different structures do not directly translate into the required sensitivity for discriminating them from the standard structure used in Expt 1. This is because each line on the graph represents data for a single rotation angle, whereas, as described earlier, the rotation angle, ϕ_2, was co-varied in the experiment with the dihedral angle of the structure.

Shown in Fig. 8 is the function describing $\Delta d/d_m$ when the rotation angle is chosen so that elements on the plane of slant θ_2 that shared the same image location in the first frame as elements on the standard structure ($\theta_1 = 60$ deg and $\phi_1 = 1$ deg) also shared the same image location in the final frame. The five particular comparison structures used in Expt 1 each occupy two points on the solid line (one for each plane), placed at values of θ_2 that are positioned symmetrically about $90 + \phi_2$ deg. The sensitivity to $\Delta d/d_m$ required to be able to discriminate each configuration from the standard (also two points on the curve at $\theta_2 = 60$ deg and 122 deg, with $\Delta d/d_m \approx 0$) can be read off against the vertical axis (note that each pair of points is positioned symmetrical around $\Delta d/d_m = 0$).

Relief SFM

In geometric terms, planarity is specified by two views. Consistent with this, the results of Expt 2 have shown that humans can reliably discriminate between planar and non-planar structures given a two-frame sequence (see also Todd & Bressan, 1990). While this precludes an analysis of the image motion requirements in terms of measurements of a single element across two displacements, an analogous analysis based on the differences in the image motion of two elements across space can be applied. Under orthographic projection, a 3-D structure that rotates about an image plane axis, is planar if the image motions of all pairs of elements symmetrically located around the rotation axis are equal. A non-planar structure can be discerned by detecting differences in the motions of such pairs of elements.

As for the recovery of metric SFM, two candidate strategies can be considered for the visual system's measurement of the image differences required to support relief SFM. One possibility is that the visual system could update the positions of image elements by measuring their location, with respect to the rotation axis, on each of the two frames. As argued above, this task can be linked to line length discrimination or element separation discrimination, both of which have Weber fractions of around 0.03. Consider two elements, one on either facet of the threshold non-planar structure, equidistant from the rotation axis in the first frame. The changes in the image positions of these elements in the second frame, as a proportion of their distance to the rotation axis, were 0.037 and 0.024. It is arguable whether this measurement technique could even support the detection of motion at all in this case. However, for this strategy to be successful, the visual system would also need to be able to discriminate the distances in the image of these two elements to the rotation axis. As the

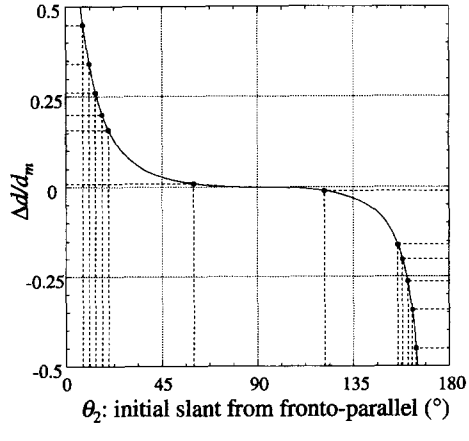

FIGURE 8. Values of $\Delta d/d_m$ for elements slanting back from the image plane by angle θ_2, and undergoing motion around a vertical axis in two rotational steps. The rotation angle, ϕ_2, is constant for a given element across displacements but varies with θ_2, such that the projected positions of all elements are identical in the first and third frame (the radial distance of the element to the rotation axis is also co-varied with θ_2). The points along the curve represent the stimuli used in Expt 1. As $\Delta d/d_m$ is constant for all elements that lie on a plane that passes through the rotation axis, each of the six stimuli are represented by a pair of such points, one for each plane. These points are symmetrical about $\Delta d/d_m = 0$ on the vertical axis and also about $90 + \phi_2$ deg on the horizontal axis. The standard structure is represented by the two points closest to $\Delta d/d_m = 0$ at $\theta_2 = 60$ and 122 deg. The five other pairs of points represent the comparison structures. Note that as the dihedral angle increases (represented by the difference between the two points along the horizontal axis) so does the magnitude of $\Delta d/d_m$.

difference in these distances, as a proportion of the mean distance, was only 0.013, it can be seen that this strategy is incapable of supporting the 3-D thresholds obtained in Expt 2.

A second strategy that the visual system might adopt would be to measure the image displacements or speeds of these two elements. An experiment by De Bruyn and Orban (1988) is relevant to investigating the effectiveness of this strategy. They measured speed discrimination thresholds for two smoothly translating random-dot patterns, seen one after the other. Under optimal conditions, Weber fractions were around 0.06, a finding replicated subsequently by Snowden and Braddick (1991).

For the threshold non-planar structure, the difference in the image displacements or speeds of two elements, placed symmetrically in the image about the rotation axis as a proportion of the mean displacement or speed, was 0.39. The fact that this value is greater than the figure obtained by De Bruyn and Orban (1988) shows that this strategy is capable of supporting the 3-D performance levels obtained in the present experiment. The following analyses of the sensitivity to image motion required to compute relief SFM assume this strategy as their basis.

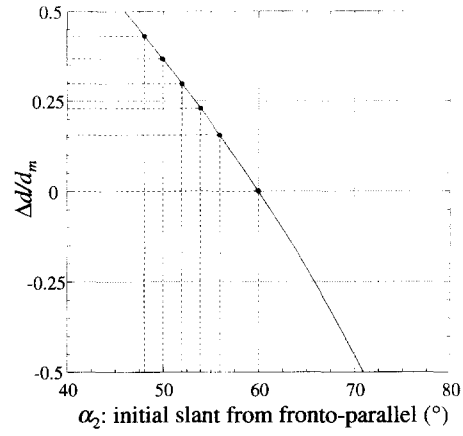

FIGURE 9. Values of $\Delta d/d_m$ derived from equation (6) a pair of elements rigidly rotating about a vertical axis. One of the elements makes an angle of 60 deg with the fronto-parallel, while the angle of the other, α_2, is variable. The rotation angle, ϕ_2, and the 3-D distance of the elements to the rotation axis are co-varied with α_2 such that the initial projections of the two elements and the sum of the two displacements are identical for all pairs of elements. Each of the structures used in Expt 2 are represented by a single point along this curve. Note that for the planar, standard structure $\alpha_2 = 60$ deg and so $\Delta d/d_m = 0$. For the comparison structures, as the deviation from planarity increases, so too does the magnitude of $\Delta d/d_m$.

Geometric analysis of planarity discrimination

Here, the consequences of errors on estimates of the image motion are considered in detail for the family of structures used in Expt 2. Note that the Weber's law process for speed discrimination studied by De Bruyn and Orban (1988) can be expressed in terms of $\Delta d/d_m$. This means that for the planarity discrimination task, the required sensitivity to image motion can also be expressed in terms of $\Delta d/d_m$. However, in terms of equation (2), d_1 and d_2 now refer to the displacements of two distinct elements, symmetrically positioned about the rotation axis, rather than to two displacements of a single element (the case for the dihedral angle discrimination task). Equation (5) describes this formulation, using the geometry and notation shown in Fig. 5(b).

$$\Delta d/d_m = \frac{[r_1 \cos(\alpha_1 + \phi_2) - r_1 \cos \alpha_1] - [r_2 \cos(\alpha_2 + \phi_2) - r_2 \cos \alpha_2]}{[r_1 \cos(\alpha_1 + \phi_2) - r_1 \cos \alpha_1] + [r_2 \cos(\alpha_2 + \phi_2) - r_2 \cos \alpha_2]/2}. \quad (5)$$

Setting the initial image positions of the two elements to be symmetrical about the rotation axis, rearranging and using trigonometric identities gives

$$\Delta d/d_m = \frac{\tan \alpha_1 - \tan \alpha_2}{[\tan \alpha_1 + \tan \alpha_2 + 2 \cos \phi_2 - 2 \operatorname{cosec} \phi_2]/2}. \quad (6)$$

When ϕ_2 is small, $\tan(\phi_2) \approx \sin(\phi_2)$ and so (6) reduces to

$$\Delta d/d_m = \frac{\tan \alpha_1 - \tan \alpha_2}{[\tan \alpha_1 + \tan \alpha_2]/2}, \quad (7)$$

such that $\Delta d/d_m$ does not depend upon the rotation angle, unlike the equivalent function for the case of metric structure described in equation (4). Equation (6) is shown graphically in Fig. 9 for the case where $\alpha_1 = 60$ deg, $\phi_1 = 1$ deg and α_2 is variable. The stimuli used in Expt 3 all lie along the curve and are denoted by the circles. The planar structure is positioned on the curve where $\alpha_2 = \alpha_1$, and so $\Delta d/d_m = 0$. As α_2 deviates from 60 deg, so the structure deviates from planarity and the magnitude of $\Delta d/d_m$ increases.

The crucial thing to note about this function is that the gradient of $\Delta d/d_m$ with α_2 is much steeper than the gradient of $\Delta d/d_m$ with θ_2, characterising the metric-structure stimuli, shown in Fig. 8. This means that a far lower sensitivity to $\Delta d/d_m$ is required in order to detect a deviation from planarity of angle α than is required to discriminate two non-planar structures that differ by that same angle. The exact ratio of sensitivity required for the two tasks will depend on the exact conditions, in particular the initial depths of the structures and the rotation angle for the metric-structure task. For the stimuli used in Expts 1 and 2, the $\Delta d/d_m$ threshold required to detect a 10 deg difference in metric structure was 0.014 whereas the $\Delta d/d_m$ threshold required to detect a 10 deg difference in relief structure was 0.37, demonstrating the much larger noise-tolerance of the relief-structure task.

RE-PLOTTING THRESHOLDS IN TERMS OF $\Delta d/d_m$

The preceding arguments make clear that any processing stage concerned with giving a 3-D interpretation to

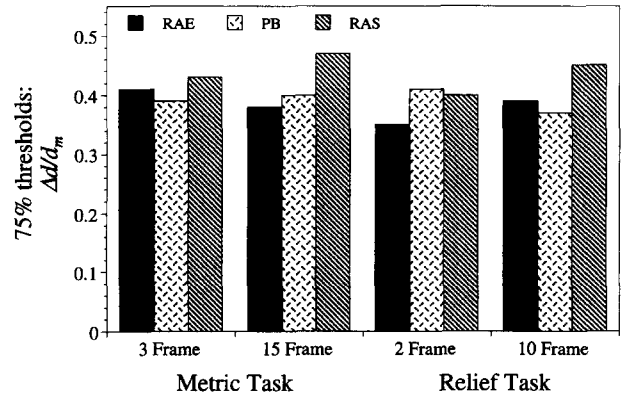

FIGURE 10. Thresholds for Expts 1 and 2 expressed in terms of the magnitude of $\Delta d/d_m$ for the comparison structures that could just be discriminated from the standard.

a particular pattern of image motion will be fundamentally constrained by the accuracy of image motion measurements. Thus, while it has been shown that in 3-D angular terms thresholds for discriminating planarity are a factor of about 8 lower than those for discriminating 3-D angles, it is important to re-plot these thresholds in terms of the associated image-motion discriminations. Figure 10 shows the values of $\Delta d/d_m$ as derived from the preceding equations for the threshold comparison structures in Expts 1 and 2. The graph reveals only negligible differences across all of the conditions—the thresholds for the data pooled across subjects and the two conditions are 0.41 for the metric-structure task and 0.40 for the relief-structure task.

How do these thresholds relate to those measured explicitly on 2-D motion tasks? Snowden and Braddick (1991) have compared speed modulation discrimination thresholds with speed discrimination thresholds. In Snowden and Braddick's study, the speed modulation task involved subjects distinguishing a random-dot pattern moving at constant velocity from one whose speed was modulated over time according to a square-wave function. In the speed discrimination task, there were two intervals both containing a random-dot pattern translating across the screen at a constant (but different) speed. Subjects were asked to judge which was faster. The modulation discrimination task is related to the present angle discrimination task (differences in displacement over time) and the speed discrimination task is related to the present planarity discrimination task (differences in displacement over space). Performance was expressed for both tasks as the Weber fraction: threshold difference in speed as a proportion of the lower speed. They found that for the speed discrimination task thresholds were as low as 0.06, whereas for the modulation discrimination task, minimum thresholds were around 0.3. Both Snowden and Braddick (1991) and Werkhoven *et al.* (1992), who found similar results, with isolated dot stimuli, suggest that the Weber fractions are poorer for the modulation detection task because observers have the additional problem of segmenting the periods of fast and slow motion, i.e. that there is uncertainty in the modulation task about the phase of the modulation.

In the present experiments, no significant differences in the thresholds were found between the metric- and relief-structures task when plotted in terms of image-motion sensitivity, which appears to be at odds with the above findings. A possible reconciliation comes from a closer inspection of how these Weber fractions vary as a function of speed. De Bruyn and Orban (1988) measured speed discrimination thresholds for a range of speeds and stimulus durations. Similarly to Snowden and Braddick (1991) they found that under optimal conditions (speeds of between 4 and 64 deg/sec, stimulus durations of at least 150 msec) Weber fractions were around 0.06. However, in one of their conditions discrimination thresholds were measured for a pattern whose speed was 1 deg/sec and whose stimulus duration was 150 msec. These settings are more in line with the motion of the outside dots in the present Expt 2. For this condition (their Fig. 8), De Bruyn and Orban found a Weber fraction of around 0.26—far closer to the thresholds recorded in the present Expt 2. Furthermore, their other data suggest that this figure would increase even more for velocities of 0.6 deg/sec—the exact speed of the outer elements in the present experiment. Snowden and Braddick (1991) also found a dependency on the temporal frequency of modulation and the absolute velocities when measuring Weber fractions for speed modulation discrimination. At the settings yielding motion similar to that in the present Expt 1 (0.48 deg/sec at 6 Hz compared to 0.61 deg/sec at 6 Hz for the outer elements) Weber fractions were around 0.35. Thus, the $\Delta d/d_m$ thresholds shown in Fig. 10 are actually only slightly greater than these 2-D thresholds.

It would be interesting to perform the SFM tasks employed in this study at a range of speeds to observe how the thresholds vary. One clear prediction is that the thresholds for the planarity discrimination task (both angular and image-motion) would decrease if rotation speed was increased.

The implication of these findings is that the sensitivity to both metric- and relief-SFM is constrained largely by the underlying sensitivity to image motion, rather than at any subsequent stages concerned with giving a 3-D interpretation to the image motion. This suggests that the visual system is *not* inherently more sensitive to relief SFM than to metric SFM. Rather, the differences in 3-D angular thresholds are due to the fact that computing metric SFM is a more noise-sensitive task.

Given this re-interpretation of the data in terms of image motion sensitivity, an important control is to show that the data do not merely reflect a response-bias in subjects to greater magnitudes of $\Delta d/d_m$, but rather show that the visual system can make more flexible use of the image motion information in interpreting 3-D structure. The following experiment aimed to investigate whether subjects could still perform the metric structure task when the magnitude of $\Delta d/d_m$ was not directly linked to the dihedral angle of the structure.

CONTROL EXPERIMENT: A VARIABLE ROTATION ANGLE CONDITION

The stimuli used for Expt 1 were designed to be amenable to a simple analysis based on sensitivity to the image-motion, where all elements on a given structure have the same magnitude of $\Delta d/d_m$. However, a possible limitation of this design is that subjects may have developed a response-strategy based on the magnitude of $\Delta d/d_m$ *per se*, rather than on an interpretation of the image motion that allowed for metric structure to be recovered under more general conditions of motion. In particular, because the comparison structure always had the larger dihedral angle, the values of $\Delta d/d_m$ for texture elements on its surface were larger than those of elements on the surface of the standard structure. In principle then, subjects could have responded, arbitrarily, on the basis of this alone. It is important to ascertain whether humans can recover metric SFM under general conditions of 3-D motion, where the magnitude of $\Delta d/d_m$ is not directly linked to the 3-D structure.

Against this hypothesis it can be noted that because there was no feedback during the experiment, it is unlikely that all three subjects would have developed this strategy, as opposed, say, to following the equally arbitrary strategy of choosing the stimulus with the smaller values of $\Delta d/d_m$ as having the greater dihedral angle. However, rather than accept this statistical evidence against the use of this strategy, a simple control experiment was performed to test this hypothesis.

Methods

The stimuli, procedure and task were largely identical to those used in Expt 1. All stimuli were shown in oscillatory three-frame motion but only two structures out of the six employed in Expt 1 were used. One was the standard ($\omega_1 = 58$ deg) and the other was one of the comparison structures ($\omega_2 = 150.4$ deg). In Expt 1, subjects were able to successfully discern that this comparison structure had a larger dihedral angle than the standard on 77% of trials.

Three conditions were run. Because one of the two observers used in this control study had not participated in the earlier experiments, one condition was simply a replication of the three-frame condition used in Expt 1.

In the other two conditions, a single variable, the rotation angle, was manipulated. For both the standard and the comparison structures, the cumulative rotation angle was kept the same as in Expt 1 (i.e. 2 deg for the standard and 13.6 deg for the comparison). This meant that the two structures still only be differed in their second-frame projections. Now though, the two steps were not necessarily kept constant. In condition 2, one of the stimuli viewed by subjects was the comparison structure rotating in two equal steps of 6.8 deg, as in Expt 1. The other stimulus was the standard structure rotating in steps of 0.78 and 1.22 deg. These angles were chosen so that the values of $\Delta d/d_m$ were 0.45 ± 0.01—similar in *magnitude* to the values of $\Delta d/d_m$ for elements on the comparison (± 0.45). If subjects were to base their responses on the magnitude of $\Delta d/d_m$, then their performance would drop to chance for this condition. In contrast, if subjects were to respond according to the correct interpretation of the image motion then performance would remain at around 75% correct.

In the third condition, subjects observed the standard structure in both intervals. In one, the rotation steps were a constant 1 deg per frame, as in Expt 1, yielding values of $\Delta d/d_m = \pm 0.01$. In the second interval, the rotations were in steps of 0.78 and 1.22 deg, as in condition 2. The prediction here is that if subjects were to choose the stimulus with larger values of $\Delta d/d_m$ as having the larger dihedral angle, then they should have a bias towards the structure in the second interval. Obviously, if subjects were to accurately interpret the structures on the basis of the image motion then no such bias would be observed.

All three conditions were tested ten times each in a block of 30 trials. Three blocks of trials were run per subject, such that each condition was tested 30 times.

Results

The data for two subjects are shown in Fig. 11. One point to note is that for the third condition, there is no "correct" response as the two structures in the trial are identical. For this case, "% correct" is used, arbitrarily, to denote the percentage of responses in which the subject chose the structure with the variable rotation steps as having the greater dihedral angle. It is clear for both conditions that subjects do not follow this response strategy. Rather, these data show that observers base their judgements of metric structure on an interpretation of the image motion that allows accurate performance under more general conditions of 3-D motion.

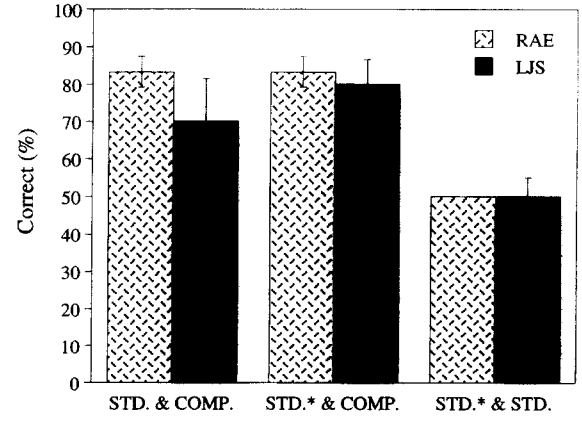

FIGURE 11. Data from two subjects performing a two-alternative forced-choice dihedral angle discrimination task. Three conditions were examined using two simulated structures. Taken from Expt 1, these were a standard (dihedral angle $\omega_1 = 58$ deg) and a comparison (dihedral angle $\omega_2 = 150.4$ deg). The stimuli each oscillated in three-frame motion. While the cumulative rotation angle of the standard was always 2 deg, and that of the comparison 13.6 deg, the steps could either be equal in magnitude (*absent) or unequal (*present).

GENERAL DISCUSSION

The central theme of this paper has been to analyse 3-D SFM tasks in terms of the required sensitivity to the 2-D image motion. Two experiments measuring discrimination thresholds for a metric and a relief SFM task have been performed. The 3-D angular thresholds were a factor of 8 higher in the metric SFM task. A model was then proposed for calculating the sensitivity to image motion required to reach particular performance levels for the two tasks. This was based on speed discrimination and speed modulation discrimination data obtained by De Bruyn and Orban (1988) and Snowden and Braddick (1991) respectively. The analysis of the 3-D discrimination performance levels in terms of this image motion model showed that thresholds were in fact similar in these terms. Furthermore, the image-motion thresholds were only marginally higher than would be expected if subjects had been asked to perform image-motion discrimination tasks. The conclusion is that processes concerned with recovering 3-D structure are not specifically tuned to relief over metric structure. Rather, recovering metric structure is inherently a more noise-sensitive process. Given these findings, it is of interest to explore how the proposed analysis applies to other empirical data on human recovery of 3-D SFM.

Metric SFM recovery

Todd and Bressan's (1990) influential paper described an experiment to investigate human recovery of metric SFM. Interestingly, they found that for both of their Euclidean-structure tasks and both of their affine-structure tasks, 3-D thresholds did not decrease as frame numbers were increased from two to eight—even though increasing the number of frames also increased the extent of the motion. Data obtained by Braunstein, Hoffman, Shapiro, Anderson and Bennett (1987) and Braunstein *et al.* (1990) suggest that frame numbers are important for determining at least some 3-D structural attributes, such as non-rigidity, although their own analysis casts some doubt on the generality of these findings. Hildreth, Grzywacz, Adelson and Inada (1990) have found that judgements of metric structure improve as frame numbers increase beyond three, but this increase in frames was confounded with an increase in total exposure duration. In considering a range of findings, Todd and Bressan (1990) concluded that as long as stimulus exposure is not confounded with the number of frames then human performance on SFM tasks does not improve beyond that achieved with two-frame sequences.

In contrast, the results of the present experiments show that the metric-structure task of 3-D angle discrimination can be performed with three-frame displays, but not with two-frame displays, even when total exposure duration is kept constant. Furthermore, although the rotation angle was larger in the three-frame condition than in the two-frame case, the symmetry of the structures about the line of sight ensured that no benefit could have been accrued from this. Higher temporal sampling, leading to a greater number of novel frames, yet with no extension of total rotation angle, led to no improvement. These findings strongly suggest that humans can make use of the additional geometrical information supplied by a third frame, when recovering 3-D metric structure.

A possible reconciliation of the present findings with those of Todd and Bressan's is that in their metric-structure tasks, the sensitivity to image motion required to make use of the additional information in the multi-frame conditions was too high. If this was so, then an improvement in performance with increasing frame numbers would not be expected. In their simplest condition, subjects were required to decide whether a two-line structure with a dihedral angle of 90 ± 22.5 deg was greater or less than 90 deg. In the present experiment, sensitivity for discriminating two dihedral angles fell short of this target by a factor of 4. Unfortunately, it is not possible to perform an analysis of Todd and Bressan's stimuli in the same terms as those used in the present study as they covaried several parameters with the 3-D dihedral angle which also affects $\Delta d/d_m$. One important parameter that they randomized was the angle at which the structures slanted back in depth. Figure 7 shows that for a rotation angle of 3 deg per frame (the value used by Todd and Bressan) $\Delta d/d_m$ varies systematically for different slants. Furthermore, unlike in the present experiment where the projections in the first and last frames were held constant for all stimuli, Todd and Bressan implemented an alternative control of systematically varying the projected line lengths, angles and displacements. Both approaches were designed to prevent any strategies based on a non-metric interpretation of the image-motion from biasing the results, but the differences may have been important, given the results of a recent study by Liter *et al.* (1994). These authors found that observers have systematic biases in perceiving metric structure based on first-order image motion properties, notably what they term relative motion (the maximal stimulus differences in image velocity). The greater this value, the more depth observers tend to perceive—regardless of its source. Thus, it may not be possible to make predictions based upon the sensitivity to $\Delta d/d_m$ even for specific trials in Todd and Bressan's task, as other factors may be at play.

The analysis presented here can also account for the results of a more recent study carried out by Norman and Todd (1993). Their stimuli simulated 3-D surfaces covered in random-dots that rotated around a vertical axis. The 3-D structure was then sinusoidally stretched and compressed over time. They found that while affine stretching distortions along the line of sight were not detected (the structures appeared to be rigid but speeding up and slowing down) distortions perpendicular to the line of sight were correctly seen as non-rigid. Norman and Todd pointed out that if observers were insensitive to the temporal relations across two displacements, both distortions would have been undetectable. They suggested that a possible explanation for their finding was that observers are sensitive to the sign of acceleration, but not to its magnitude. In fact, this result

is also compatible with the suggestion of the present paper that humans *do* have a measurable sensitivity to the magnitude of $\Delta d/d_m$, but that it is quite poor. This is so because the image motion produced by even quite large non-rigid distortions along the line of sight is very similar (though not identical) to that produced by a rigid structure rotating at a variable angular velocity (Pollick, 1995), meaning that high sensitivity is required to discriminate them.

Future work

The model developed in this paper to account for SFM thresholds generates several other testable predictions. One is about the relative dependencies of 3-D angular thresholds for the metric- and relief-structure discrimination tasks on the rotation angle used. The metric-structure task is predicted to be dependent on the rotation angle, whereas the relief-structure task is predicted not to be [equations (4) and (6)]. Intuitively, as the analysis for the metric-structure task is based on the differences in image motion for single elements across multiple displacements, the magnitude of $\Delta d/d_m$ will clearly be dependent on the extent of the motion. For the relief-structure task, because the analysis is based upon the ratio of the image motions of two different elements, $\Delta d/d_m$ will depend almost exclusively on the different positions of these elements in 3-D space. Performing these tasks for a range of rotation angles would be a simple test of these hypotheses.

A second prediction lies at the heart of the argument that 3-D SFM is not directly recovered by the visual system but is in fact derived from and highly constrained by lower-level image-motion processing. If 3-D SFM recovery is a fundamental process carried out by the visual system then angular discrimination thresholds should not vary as a function of the 3-D structures used and their initial orientations in 3-D space. However, if 3-D thresholds are constrained by 2-D thresholds then systematic variations in thresholds would be predicted for different 3-D structures. In particular, for the type of bi-planar stimuli used in Expt 1, discrimination thresholds are predicted to be lower when the dihedral angle of the standard is greater, i.e. when the planar facets are angled closer to the image plane. Figure 7 shows that the gradient of $\Delta d/d_m$ with respect to slant is steeper when the facets are slanted close to the image plane, illustrating that mis-estimations in $\Delta d/d_m$ would lead to relatively small errors in the estimations of dihedral angle. It should be noted that the threshold would never actually approach zero because the absolute value of Δd becomes too small to detect.

To increase the generality of these findings, it would be of great interest to measure thresholds for similar discrimination tasks for structures that rotated around axes not in the image plane. The image motion of a single feature in 3-D space traces out an elliptical path with the aspect ratio being equal to the sine of the slant of the rotation axis back from the image plane. The extraction of information about metric structure, in this case, requires sensitivity to the changes in the *direction* of image elements over time as well as to the changes in displacement magnitude. Werkhoven *et al.* (1992) have already measured 2-D motion thresholds for single elements whose direction of motion is sinusoidally modulated over time and the data have revealed a Weber's law process, at least for modulation frequencies below about 1 Hz. In general, it might be expected that 3-D thresholds for such a task would be lower than those reported here because rotations around an axis out of the image plane contain a rotation component about the line of sight, which does not contain any information about 3-D structure.

REFERENCES

Bradshaw, M. F. (1989). The combination of stereo and motion information in human vision. Unpublished Ph.D. thesis, University of Sheffield, England.

Braunstein, M. L., Hoffman, D. D. & Pollick, F. E. (1990). Discriminating rigid from non-rigid motion: Minimum points and views. *Perception & Psychophysics, 47*, 205–214.

Braunstein, M. L., Hoffman, D. D., Sharpiro, L. R., Anderson, G. J. & Bennett, B. M. (1987). Minimum points and views for the recovery of three-dimensional structure. *Journal of Experimental Psychology: Human Perception and Performance, 13*, 335–343.

De Bruyn, B. & Orban, G. A. (1988). Human velocity and direction discrimination measured with random-dot patterns. *Vision Research, 28*, 1323–1335.

Engen, T. (1972). Psychophysics: I. Discrimination and detection. In Kling, J. W. & Riggs, L. A. (Eds), *Woodworth and Schlosberg's experimental psychology* (3rd edn). London: Methuen.

Foley, J. D., van Dam, A., Feiner, S. K. & Hughes, J. F. (1990). *Computer graphics: Principles and practice* (2nd edn). Reading, Mass.: Addison-Wesley.

Gibson, E. J. & Gibson, J. J. (1957). Continuous perspective transformations and the perception of rigid motion. *Journal of Experimental Psychology, 54*, 129–138.

Grzywacz, N. M. & Hildreth, E. C. (1987). Incremental rigidity scheme for recovering structure from motion: Position-based versus velocity-based formulations. *Journal of the Optical Society of America A, 4*, 503–518.

Hildreth, E. C., Grzywacz, N. M., Adelson, E. H. & Inada, V. K. (1990). The perceptual buildup of three-dimensional structure from motion. *Perception & Psychophysics, 48*, 19–36.

Hoffman, D. D. & Bennett, B. M. (1986). The computation of structure from fixed-axis motion: Rigid structures. *Biological Cybernetics, 54*, 71–83.

Hogervorst, M. A., Kappers, A. M. L. & Koenderink, J. J. (1993). Perception of metric depth from motion parallax. *Perception (Suppl.), 22*, 101.

Johnston, E. B., Cumming, B. C. & Landy, M. S. (1994). Integration of stereopsis and motion shape cues. *Vision Research, 34*, 2259–2275.

Koenderink, J. J. & van Doorn, A. J. (1975). Invariant properties of the motion parallax field due to the movement of rigid bodies relative to the observer. *Optica Acta, 22*, 773–791.

Koenderink, J. J. & van Doorn, A. J. (1991). Affine structure from motion. *Journal of the Optical Society of America A, 8*, 377–385.

Liter, J. C., Braunstein, M. L. & Hoffman, D. D. (1994). Inferring structure from motion in two-view and multiview displays. *Perception, 22*, 1441–1465.

Longuet-Higgins, H. C. & Prazdny, K. (1980). The interpretation of moving retinal images. *Proceedings of the Royal Society of London B, 208*, 385–397.

Norman, J. F. & Todd, J. T. (1993). The perceptual analysis of structure from motion for rotating objects undergoing affine stretching transformations. *Perception & Psychophysics, 53*, 279–291.

Pollick, F. E. (1995). The perception of motion and structure in structure-from-motion: Comparisons of affine and euclidean formulation. Submitted to *Journal of Experimental Psychology*.

Rogers, B. J. & Graham, M. E. (1979). Motion parallax as an independent cue for depth perception. *Perception, 8*, 125–134.

Snowden, R. J. & Braddick, O. J. (1991). The temporal integration and resolution of velocity signals. *Vision Research, 31*, 907–914.

Todd, J. T. & Bressan, P. (1990). The perception of 3-dimensional affine structure from minimal apparent motion sequences. *Perception & Psychophysics, 48*, 419–430.

Ullman, S. (1979). *The interpretation of visual motion*. Cambridge, Mass.: MIT Press.

Ullman, S. (1983). Recent computational studies in the interpretation of structure form motion. In Beck, J. & Rosenfeld, A. (Eds), *Human and machine vision*. New York: Academic Press.

Ullman, S. (1984). Maximizing rigidity: The incremental recovery of 3-D structure from rigid and non-rigid motion. *Perception, 13*, 255–274.

Wallach, H. & O'Connell, D. (1953). The kinetic depth effect. *Journal of Experimental Psychology, 45*, 205–217.

Watt, R. J. (1987). Scanning from coarse to fine spatial scales in the human visual system after the onset of a stimulus. *Journal of the Optical Society of America A, 4*, 2006–2021.

Weibull, W. A. (1951). A statistical distribution function of wide applicability. *Journal of Applied Mechanics, 18*, 292–297.

Werkhoven, P., Snippe, H. P. & Toet, A. (1992). Visual processing of optic acceleration. *Vision Research, 32*, 2313–2329.

Westheimer, G. & McKee, S. P. (1979). What prior uniocular processing is necessary for stereopsis? *Investigative Ophthalmology and Visual Science, 18*, 614–621.

Acknowledgements—This work was supported by an EC ESPRIT grant No. 6019. We are grateful for the support and encouragement of Brian Rogers, and for helpful discussions with Mark Bradshaw, Andrew Glennerster and Jim Todd.